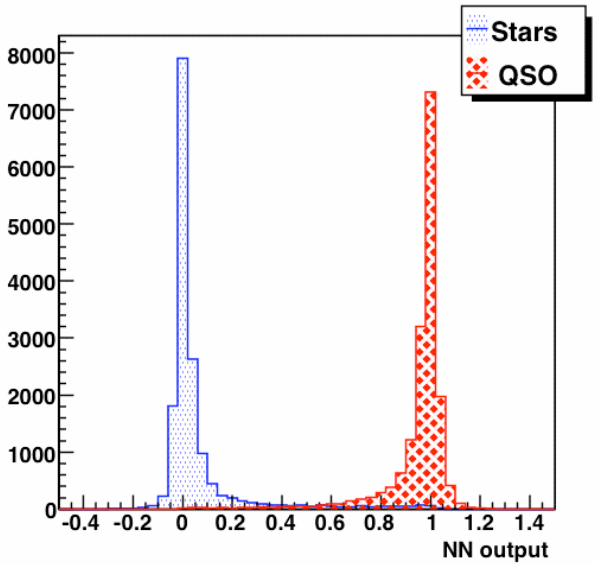
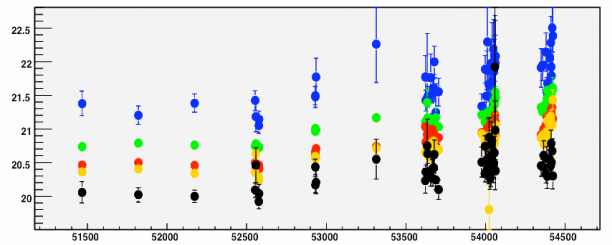
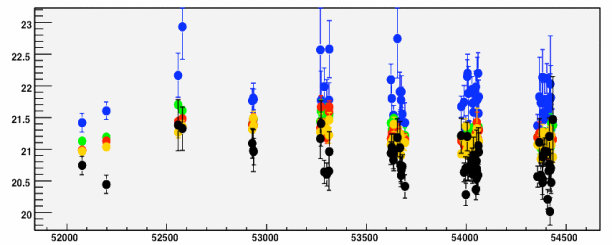
QSO Selection with NN

Ch.Yèche (CEA-Saclay)
on behalf of BOSS FPG







Outline:

QSO Selection with photometry

- Star/QSO separation
- Redshift
- Performances for BOSS

QSO Selection with variability

- Performances for BigBOSS

BigBoss Collaboration Meeting, Berkeley, Nov. 18-20, 2009

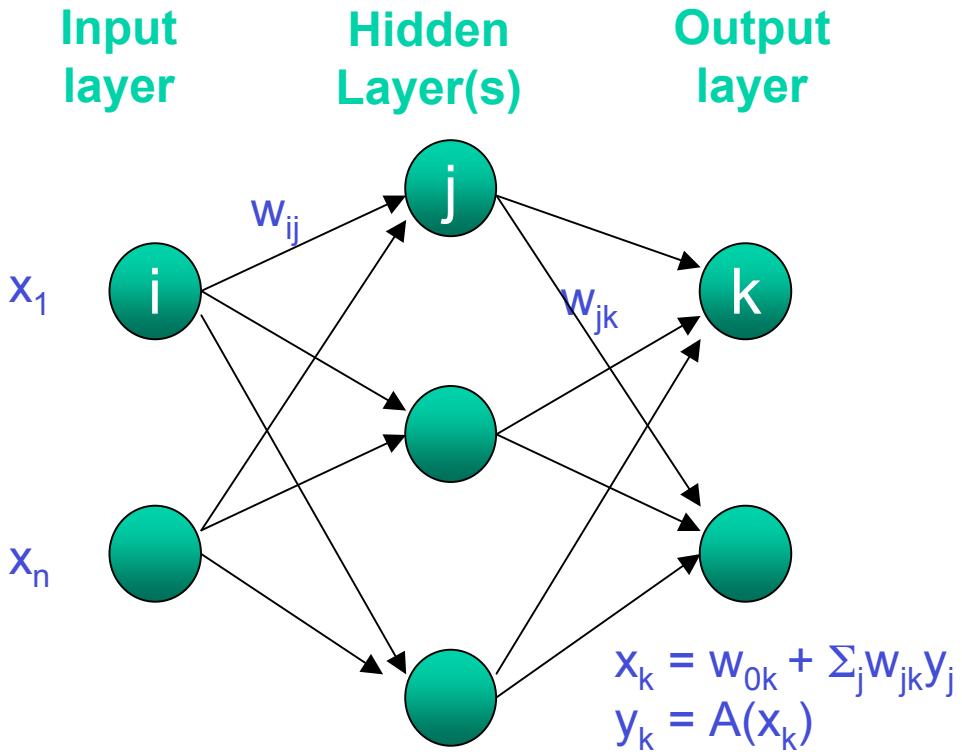
Strategy for QSOs/Stars separation

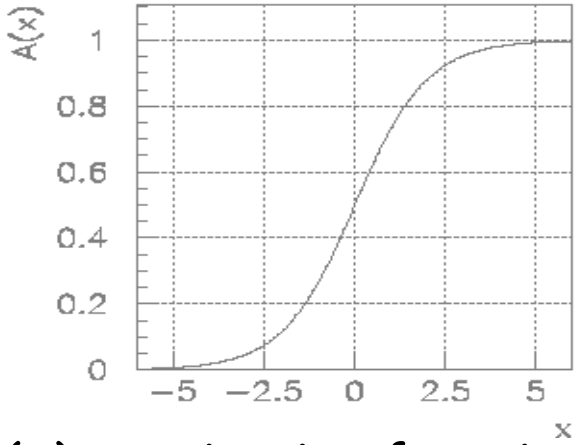
- **QSO selection in BOSS**
  - ⇒ For BAO at least 15 QSOs per deg^2
  - ⇒ Completeness : 20-25 /deg^2 with $g<22$ and $2.2<z<3.5$
  - ⇒ Many more stars than QSOs (x 200-500)
  - ⇒ Single Epoch photometry (but stripe 82)

- **NN algorithm**
  - ⇒ Alternative approach to two other methods (KDE, likelihood)
  - ⇒ Use more variables and correlation between them
  - ⇒ Compute both star/QSO separation and QSO redshift

- See paper : *Ch. Yèche et al., arXiv:0910.3770v1*

Discrimination with a NN





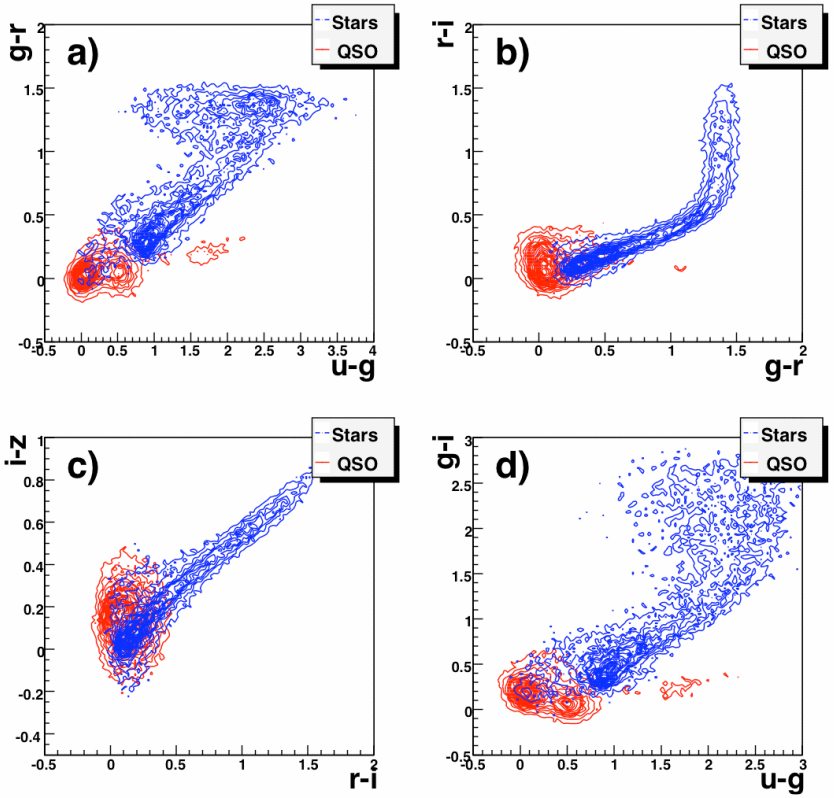
A(x) : activation function, sigmoid function ($1/(1+e^{-x})$, tanh(x)…)

Principles:

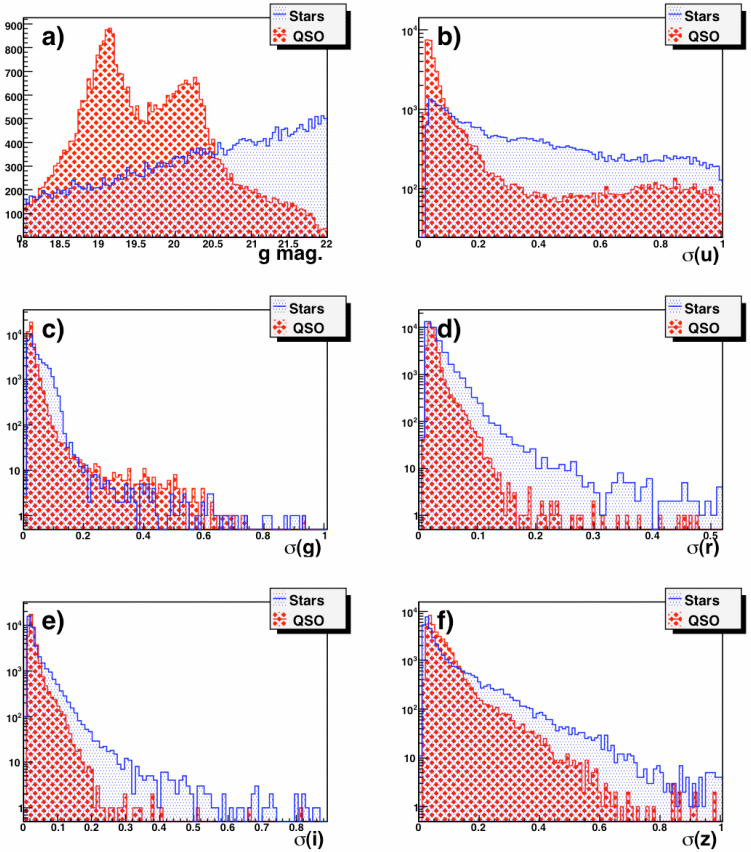
- Connection between the **nodes (neurons)** with a smooth function A(x)
- In practice, we need **training samples (Stars and QSOs)** to compute the weights w_{ij} by minimization of a pseudo-χ^2
- One output node y_0: Learning with **y_0=0 if stars ; y_0=1 if QSOs**

QSO Selection with NN

Based on SDSS-DR7 imaging database



4 colors (u-g,g-r,r-i,i-z)



Add 6 variables:

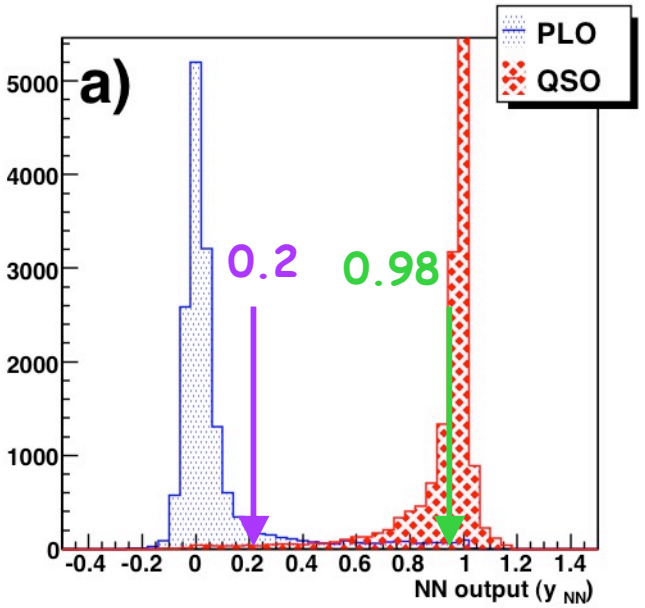
g and 5 magnitude errors

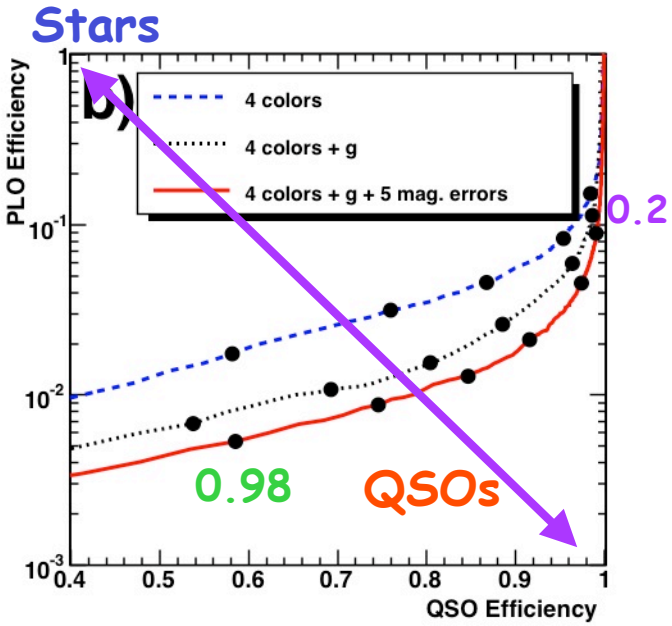
σ(u), σ(g), σ(r), σ(i), σ(z)

QSO Selection with NN

- NN 4 Var: 4 colors
- NN 5 Vars: 4 colors +g
- NN 10 Vars: 4 colors +g + magnitude errors

⇒ Better performances with 10 variables

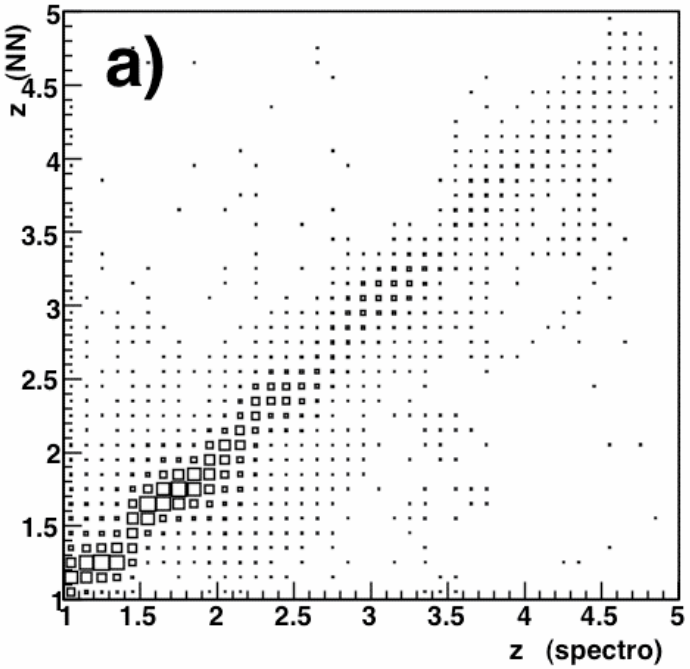


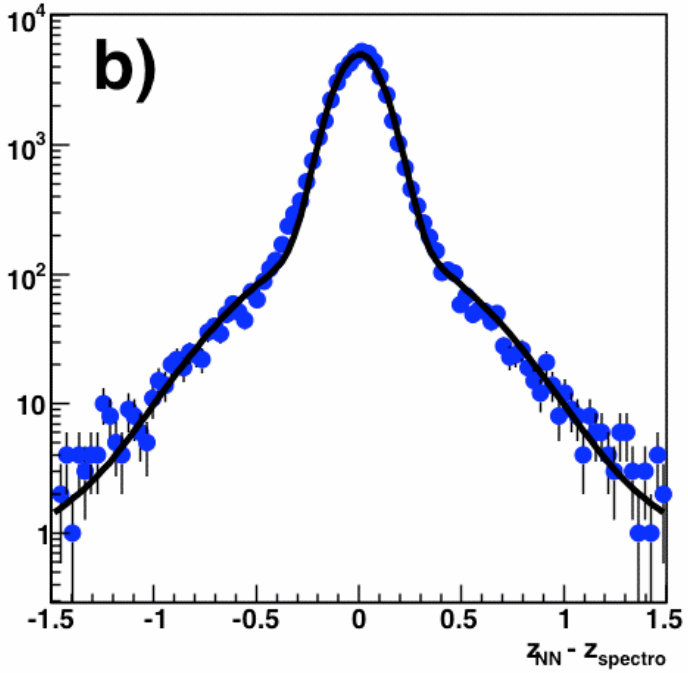


- QSO efficiency 50% --- Stars contamination 0.4%
- 15 QSOs/deg² for a fiber budget of ~30/deg²
- Validation of this method with commissioning data for fainter objects…

QSO Photo z with NN

- We can do better by using the z as NN output!!!
- Prediction of the z with the NN ⇒ **Photo-z**
- Crucial to select interesting QSOs for BAO



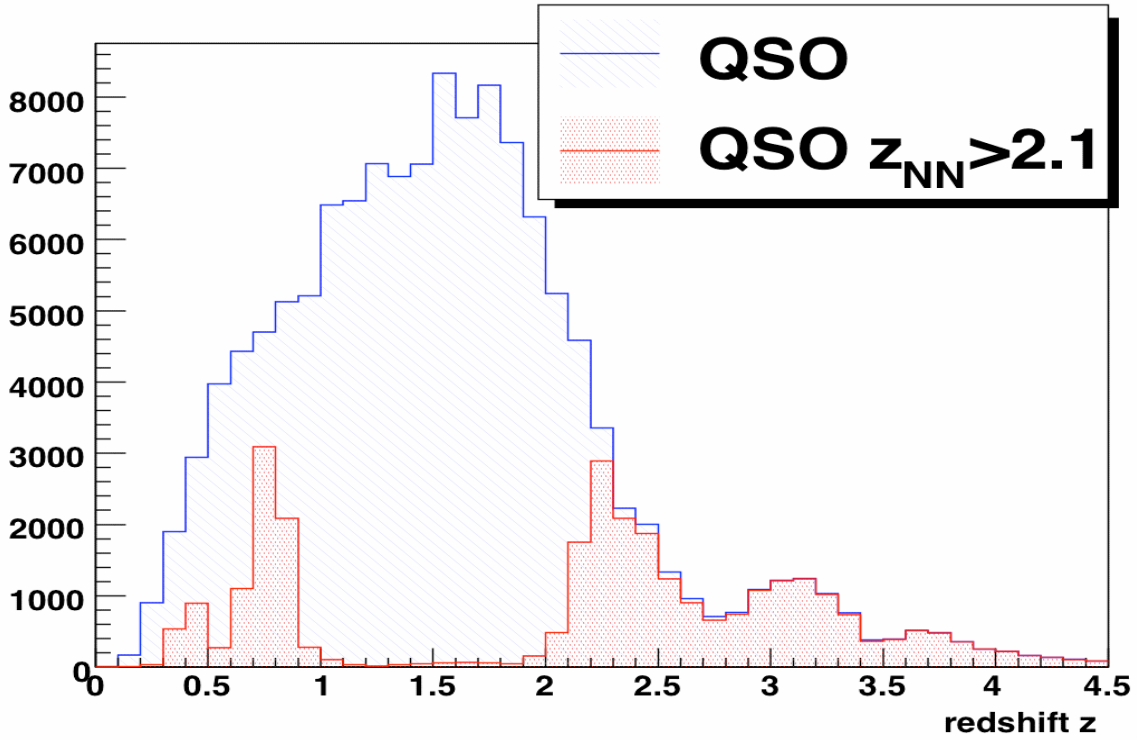


3 Gaussians:

- Core f=93.4%
  σ=0.1
- Wide f=6.4%
  σ=0.4
- Outliers f=0.2%
  σ=1.0

Select QSOs with z>2.2

- To get rid of QSOs below 2.2, we applied a cut z_{NN} >2.1



- 90% of the QSOs with z<2.2 are removed
- 5% of the QSO with z>2.2 are lost

Figure of Merit for NN

- Selection of the "best" QSOs for BAO studies
- QSO Probability

$$P_{QSO}(x_{NN},\alpha,\delta) = \frac{N_{QSO} f_{QSO}(x_{NN})}{N_{QSO} f_{QSO}(x_{NN}) + N_{Star}(\alpha,\delta) f_{Star}(x_{NN})}$$

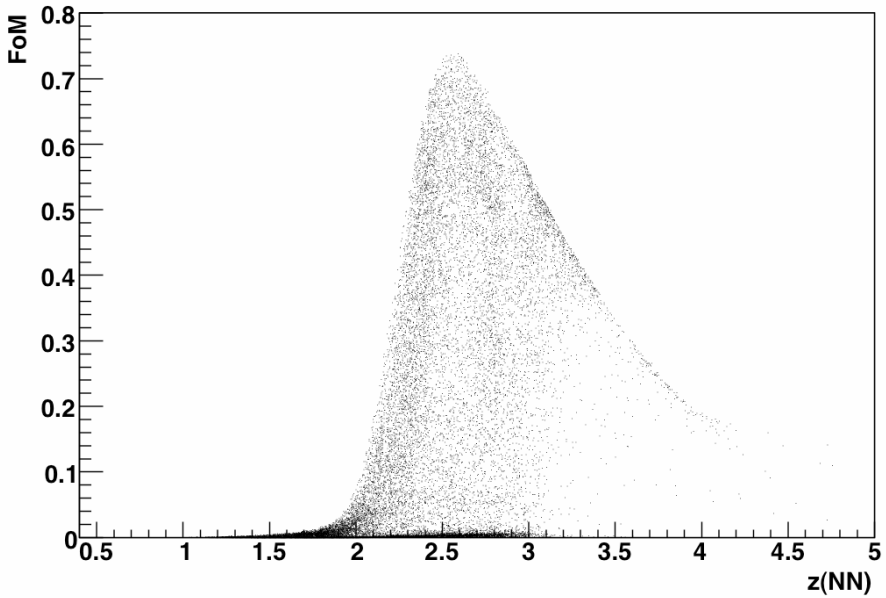
- Pat McDonald's FoM : V(z,g)

Convolution with
resolution of photo-z

$\Rightarrow$ $\tilde{V}(z_{NN},g)$

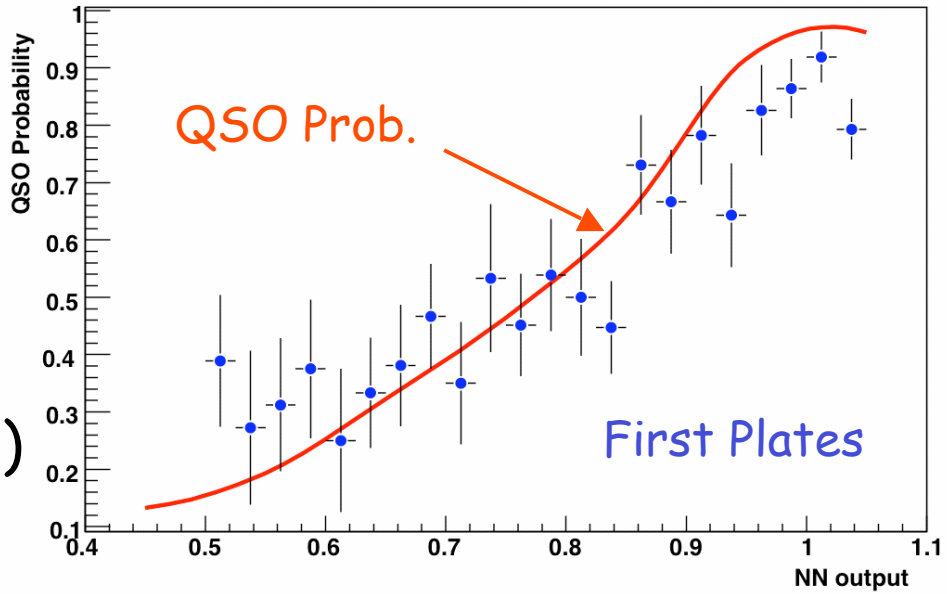
- NN FoM

$$FoM_{NN}(x_{NN},\alpha,\delta,z_{NN},g) = P_{QSO}(x_{NN},\alpha,\delta)\cdot\tilde{V}(z_{NN},g)$$



First Results with Commissioning Data

- **Equatorial SDSS Stripe 82**
  - ⇒ Co-Add Photometry (3 methods: 80 obj/deg^2)
  - ⇒ Completeness : ~23 QSO/deg^2
  - ⇒ For NN ~20 selected obj/deg^2
  - ⇒ Purity (Eff.): 67% QSO
    58% QSO (z>2.2)



- **Single Epoch Photometry**
  - ⇒ For NN ~20 selected obj/deg^2
  - ⇒ Purity : 37% QSO, 32% QSO (z>2.2)
  - ⇒ **Retraining required** for fainter QSO (g>21) to achieve 40-50%

- ***A guess:*** Very difficult to obtain a purity>50% with single epoch photometry, even by combining several methods…

Improve QSO selection with Variability

Statement in BigBOSS white paper:

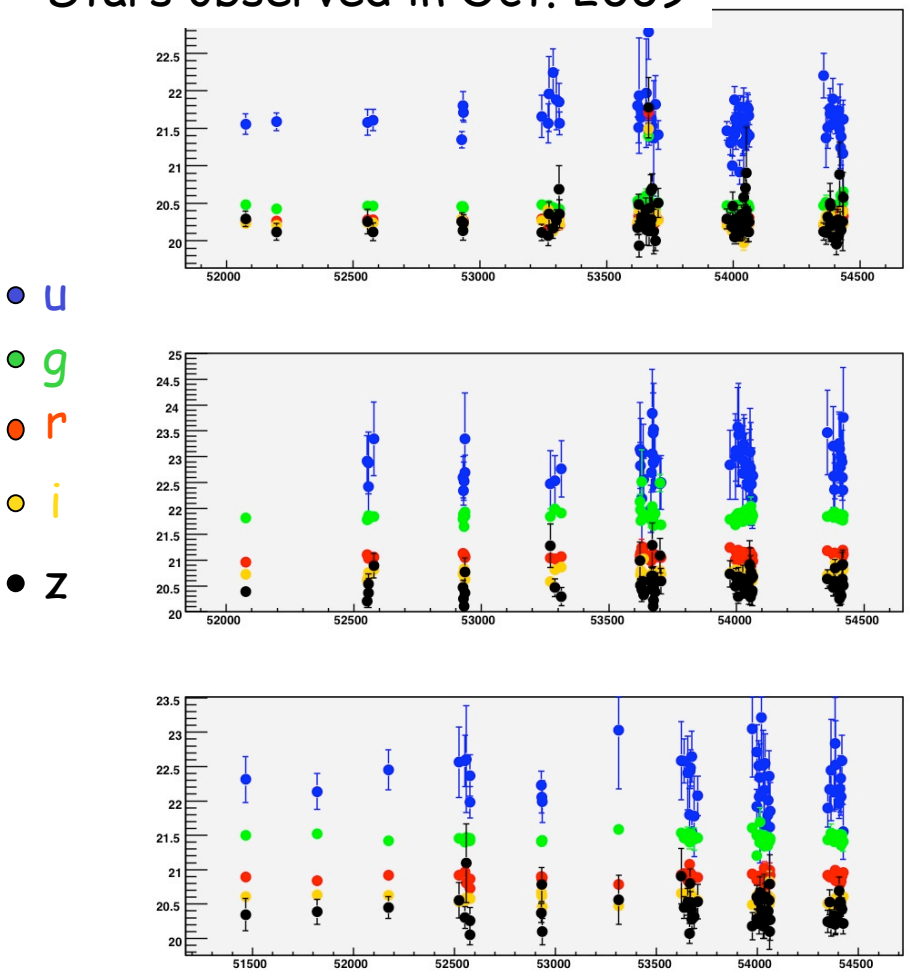
"The selection of QSOs is straightforward from either color selection or variability. BOSS selects QSOs in the redshift range 2<z<3.5 with an efficiency of 40% using SDSS colors to g=22. This efficiency can be increased with u-band data deeper than SDSS. However, variability from Pan-STARRS and other imaging surveys will allow the selection of QSOs with a completeness exceeding 85%"

Test of the method:

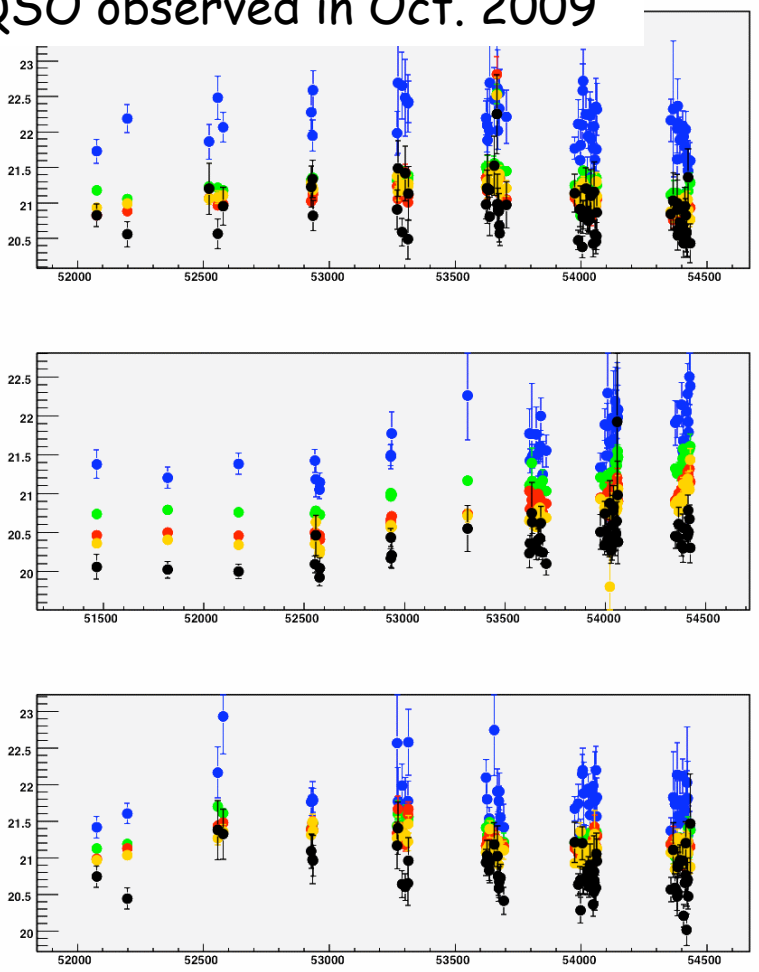
- SDSS stripe 82 (observations over 7 years)
- Use QSO and Stars spectroscopically confirmed by BOSS selected with NN as QSO on Stripe 82
- Light curves for these objects
- Development of a new NN to select QSOs

Commissioning: Stripe 82 Light Curves

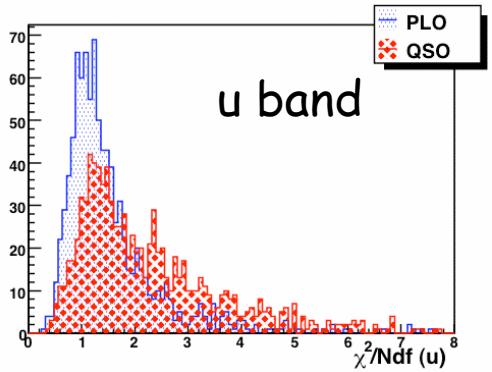
Stars observed in Oct. 2009

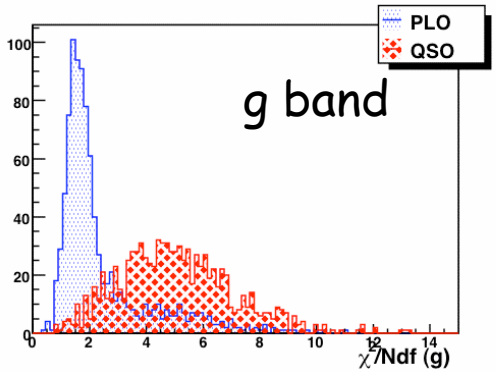


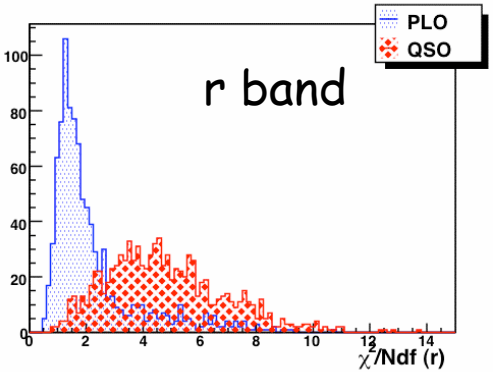
QSO observed in Oct. 2009

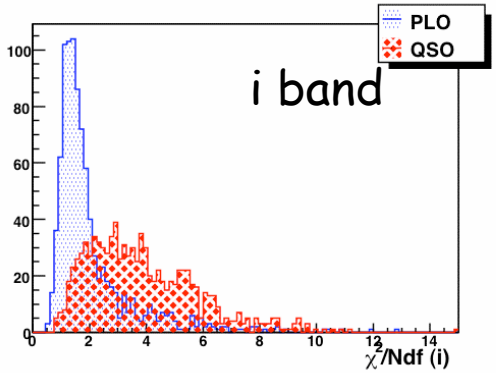


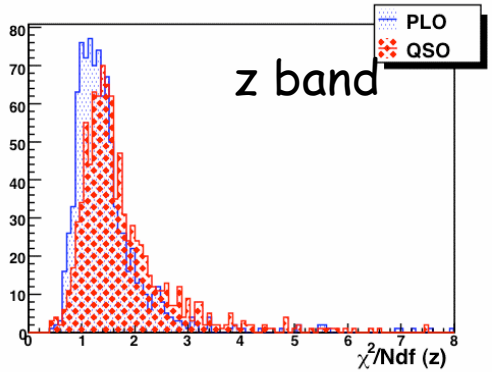
QSO variability





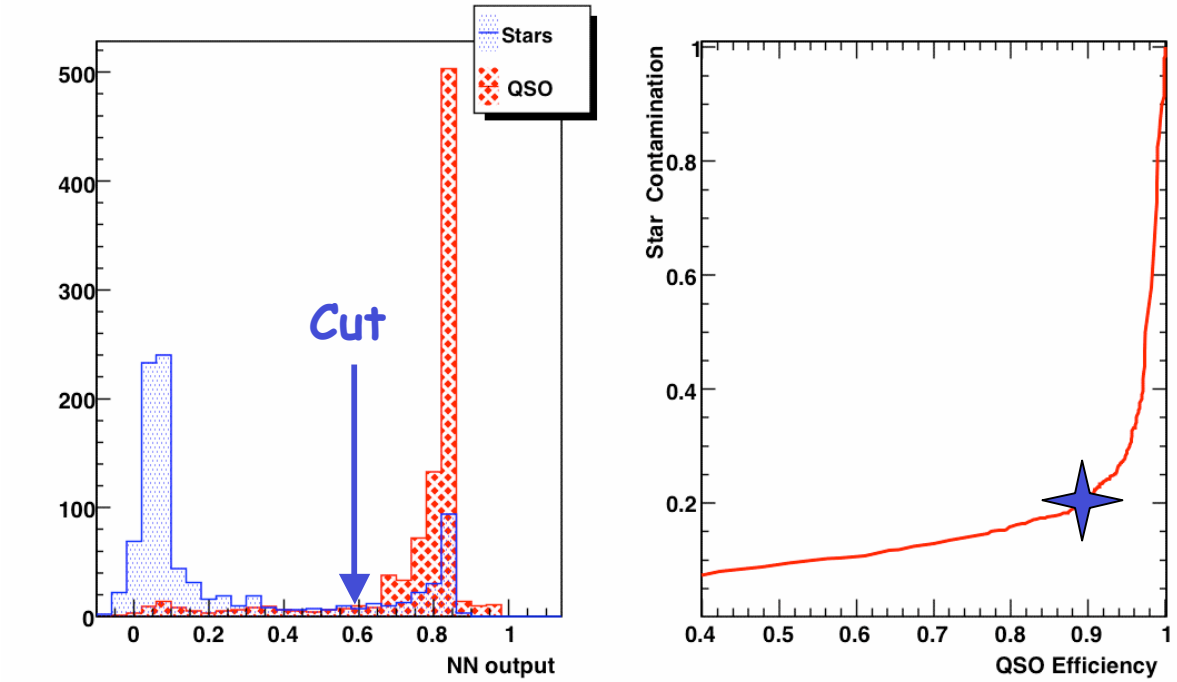






- **First attempt:** Simple and robust variables, much more sophisticated variables can be used!!!
- χ^2 **value** of a constant fit after 5-σ clipping for the five bands *ugriz*
- Main information from the 3 bands *gri*
- Instead of applying cuts on variables ⇒ **Combine the 5 variables with a NN** (see next slide)

NN for QSO variability



- NN combines the 5 variables
- Training with objects of stripe 82 which pass the photometry NN

- A 90% QSO efficiency for a 80% star rejection

- Combining the two NNs (photometry and variability) :

25 fiber budget : 20 QSOs and 5 stars for g<22 (total purity ~80%)

Summary

- New approach with NN to select QSOs against stars and to compute photometric redshit
- Preliminary results look promising but need to be confirmed with first BOSS data
- Retraining for fainter QSOs is very likely required
- Adding variability will allows us to achieve a 80% purity for BigBOSS